iSAM: Personalizing an Artificial Intelligence Model for Emotion with Pleasure-Arousal-Dominance in Immersive Virtual Reality

Conference Paper · January 2020

DOI: 10.1109/FG47880.2020.00091

Aviv Elor
University of California, Santa Cruz

iSAM: Personalizing an Artificial Intelligence Model for Emotion with Pleasure-Arousal-Dominance in Immersive Virtual Reality

Aviv Elor and Asiiah Song
University of California - Santa Cruz, Department of Computational Media, Santa Cruz, California, USA
{aelor, julinas}@ucsc.edu

Abstract— **Emotion, a crucial element of mental health, is not often explored in the field of immersive Virtual Reality (iVR). Enabling personalized affective iVR experiences may be incredibly useful for the expansion and evaluation of serious games. To further this direction of research, we present a playable iVR experience in which the user evaluates the emotion of images through an immersive Self-Assessment Manikin (iSAM). This game explores a pilot system for enabling efficient online fine-tuning of a user's Pleasure-Arousal-Dominance (PAD) emotional model using personalized deep-learning. We discuss adapting the International Affective Picture system (IAPs), in which our Artificial Intelligence (AI) model responds with a personalized image after learning from ten user supplied answers during an iVR session. Lastly, we evaluated our iVR experience with an initial pilot study of four users. Our preliminary results suggest that iSAM can successfully learn from user affect to better predict a 'happy' personalized image than the static base model.**

I. Introduction

At this time, iVR Head Mounted Display systems (HMDs) have garnered wide commercial adoption, with over 200 million projected headsets sold since 2016 [1]. In addition to entertainment purposes, iVR holds vast potential for serious games, which has been on the rise due to the benefits of programmable iVR for physical and cognitive applications [2]. High-fidelity motion capture and telepresence capabilities allow these assitive experiences performed in VR environments to increase user game compliance, accessibility, and data throughput while using commercially available components [3], [4], [5], [6]. In this paper, we explore how iVR can be utilized to personalize an experience for emotion using the Unity3D Game Engine and the HTC Vive iVR System. The system personalization was all performed during runtime to create a unique situation for every user. This research may be of interest to interdisciplinary researchers at the intersection of immersive media, artificial intelligence, and healthcare intervention.

In terms of modeling emotion, Paul Ekman describes nine principles for basic emotions. Ekman argues that: emotions have universal signals, are found between animals, affect physiological systems, are triggered by universal events, are coherent, have rapid onset, brief duration, are appraised automatically and subconsciously, and they are involuntary [7]. These principles provide a theoretical framework for quantifying emotions and starting empirical studies on affective states. Subsequently, many researchers have explored how to quantify these basic emotions (anger, fear, sadness, enjoyment, disgust, and surprise) in media such as music and photos [8], [9]. Considering these kinds of works, how might an emotional state be directly mapped through an iVR environment that can also provide coherent and fast responses to user interactions?

To answer this question we chose to utilize the Pleasure-Arousal-Dominance (PAD) emotional model, a conceptual construct explaining that human responses to environments can be quantified in terms of three independent bipolar dimensions [10], [11]. These dimensions of PAD can describe the emotional response from environments through pleasure-unpleasant (P), arousal-unaroused (A), and dominant-submissive (D) states[11]. To measure PAD from our users in an iVR environment, our project uses the Self-Assessment Manikin (SAM), one of the most widely used surveys for evaluation of emotional states. SAM allows for quick, non-verbal, culture-free, and language-free retrieval of PAD response to a given stimuli [12]. Applications of SAM include the University of Florida's Center for the Study of Emotion and Attention (CSEA) affective databases for pictures, audio, and words [13]. The CSEA database contains statistically-based media-to-affective-value models and has been explored with Event Related Potentials, functional Magnetic Resonance Imaging, Pupil Dilation, and more [14], [15], [16], [17]. For example, Waltemate et al's avatar personalization study utilized SAM to evaluate emotional and social experience in response to presence and immersion in embedded user avatars in an iVR environment [18].

In terms of creating adaptive experiences with PAD evaluation, i Badia et al. employed biofeedback to infer affect using the International Affective Picture system (IAPs [13]) while users navigated a virtual maze [19]. This study employed a post-test SAM, but did not query subjective feedback of self-perception from users during gameplay. With considerations of the works discussed in these sections, we sought to create an experience that could learn from the user's PAD response in a runtime iVR game. To this end, we decided to employ SAM as a runtime input mechanic to enable appraisals of user emotional states during an interaction, this then informs the prediction of our AI. Translating the PAD emotional model into a runtime mechanic for game engines may yield immense potential in an iVR environment. This work will lead to a model that can help personalize emotional engagement which could lead to more effective experiences for a variety of serious game applications in health, rehabilitation, and entertainment [2].

II. System Design

This project, which we dub iSAM (the immersive Self Assessment-Manikin), leverages the capabilities of immersive VR and AI to create a playable experience that learns from the emotional response of players. Specifically, there were three tasks:

- Incorporate the SAM Pleasure-Arousal-Dominance emotional model into an iVR experience [15], [14].
- Establish a methodology of dynamic learning from user emotional responses for adaptive affective models.
- Evaluate the effectiveness of said methodology in dynamically adapting affective models.

We explored several approaches in our prototype to understand how emotion may be factored into curated iVR stimuli, such as a rule-based heuristic model, a random forest model, and a ResNet-based deep learning model. The prototype uses images from the IAPs, but we structured it to allow extension into domains beyond images – for example, sound or aromatics. We log user responses during runtime for evaluation and potential reuse in other projects personalizing iVR stimuli through emotion. The potential of constructing new affective databases could be useful, especially in domains where the IAPs are lacking, such as aromatics, 3d audio, and haptic feedback.

A. Interaction Design

In our iVR experience, the user enters a room which is semi-enclosed and floats in an empty neutral space. A picture frame is in front of them. The user is told that the picture within the picture frame contains a "lost memory" of a virtual character, "Sam." By responding to each of the images presented to them using the SAM affective rating scale [14], they are helping Sam recover a lost happy memory. The user inspects an image in the picture frame for 12 seconds, after which a screen appears on their left wrist, showing a SAM affective rating scale. The user evaluates the image and submits their PAD response. After this, the picture frame updates to a new image, which the user again inspects for 12 seconds before rating. The interface used by the evaluator is illustrated on the left side of Figure 2. The user's view of the SAM interface and a third-person view of the iSAM environment are shown on the right side of Figure 2.

Each of the responses of the player allows the AI environment to update its model of image-to-affect corresponding from the baseline IAPs-based model to a more personalized responses. After appraising ten pictures from the IAPs, we show the player one last image, which we present as a lost and rediscovered happy memory of Sam. The flow of the interaction for the user is shown in Figure 1, and the underlying logic driving this experience can be seen in Figure 3.

B. Implementation

We designed iSAM using several components: the Unity3D Game Engine, a python server for affective models including machine learning models, and two databases: the IAPs database and our logfile system database in which

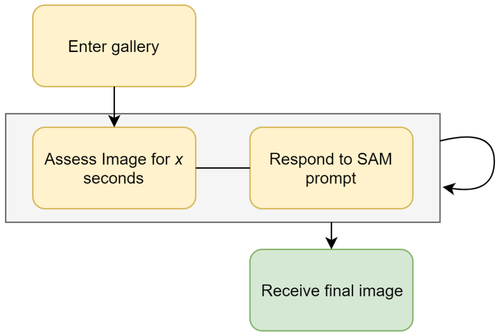

Fig. 1. iSAM user interaction flow.

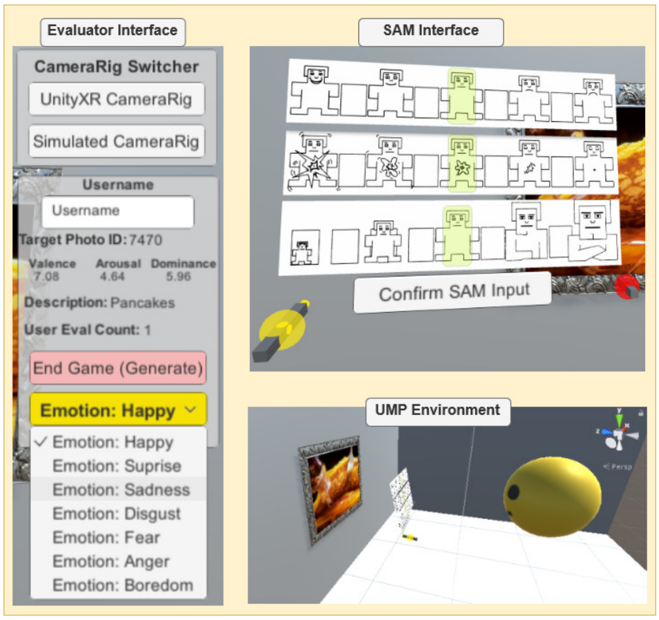

Fig. 2. iSAM evaluator interface (left), Input Interface (top right), and Gameplay Environment (bottom right).

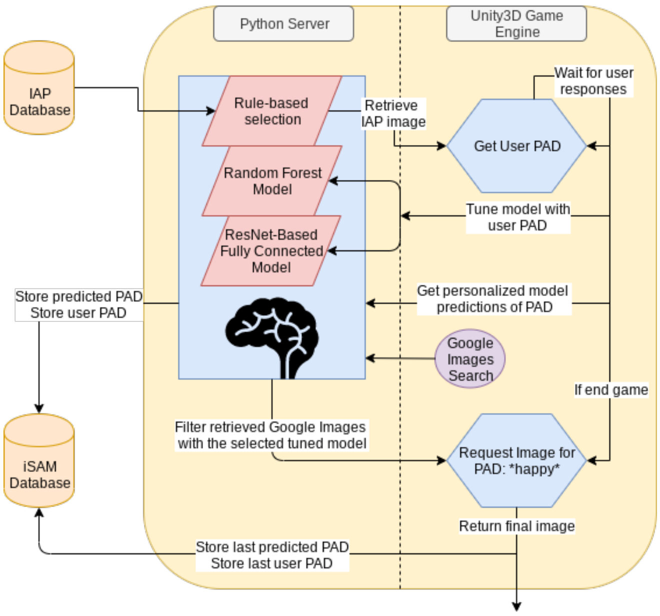

Fig. 3. iSAM logical flow

Emotion Type	P	A	D
Happy	9	5	5
Surprise	9	9	5
Sadness	1	1	9
Disgust	1	1	1
Fear	1	9	9
Anger	5	9	1
Neutral	5	5	5

TABLE I

PAD INFERENCE ASSUMPTIONS FOR THE ISAM EXTREME EMOTIONAL STATES. THE SCALE IN SELF-ASSESSMENT-MANIKIN IS FROM 1 TO 9 [20], [21], [22], [23]. THE ASSUMPTIONS ABOVE WERE USED TO RETRIEVE IAPS REFERENCES WITH A 1F DELTA RANDOM RANGE. IAPS IMAGES ARE NON-REPEATING. FINAL CURATED IMAGES WERE PRODUCED BY EVALUATING USER CSV FILES TO SEARCH FOR MATCHING PAD ASSUMPTIONS. THE PREDICTED EMOTIONAL STATE OF THE SUBJECT AND THE PAD ASSUMPTIONS WERE FED INTO THE AI MODEL TO GET THE BEST MATCHING "HAPPY" IMAGE. (P $\rightarrow$ 1 UNPLEASANT TO 9 = PLEASANT, A $\rightarrow$ 1 CALM TO 9 EXCITED, AND D $\rightarrow$ 1 INDEPENDENT TO 9 DEPENDENT [14])

user data was stored during runtime ("iSAM database" in Figure 3). We utilize these databases to handle the retrieval of IAPs baseline PAD data into the iSAM experience, the retrieval of runtime user responses, user response logging, and user emotional model inference. For each round of user interaction, the "Get USER PAD" component retrieves an image from the IAPs database and waits for user response. The IAPs database consists of pairs of images and metadata, which includes image category, as well as the pleasure, arousal, and dominance factors of each image with the standard deviation for each value. There are three versions of baseline metadata from the IAPs database: one with data from female participants, one from male participants, and one from all participants combined [13]. For our pilot test, we used the metadata version from all participants – because our user pool was so small, it did not make sense to segment by gender.

Two control methods were designed for both the environment and player input. Environment control is enabled through an IAPs picture frame in the Unity Scene, as shown in Figure 4, and handles the updates and communication to the IAP database. We implemented the frame through the Unity3D's *Texture2D* class attached to a game object allowing us to switch images in and out of the picture frame. Secondly, user input control was curated through a 3DUI representing SAM. With simplicity in mind, we used trigger colliders on the VR controller to enable button presses on a world level canvas. Thus, iSAM can be played by any VR device that has a controller, and other affective survey formats can be easily translated into the game environment by directly importing sprites. The VR controller is represented by a rectangular pointer, as seen with the yellow sphere in Figure 2 near the SAM interface. A button click is done by passing the VR controller through the desired PAD value. Once a valid PAD value is given to every dimension, the "Confirm SAM Input" button becomes available for the user to store their response. Subsequently, this confirmation builds a string with a global timestamp that holds response time and PAD value to be stored and communicated in CSV format by the python server.

As user PAD responses come in, they are sent to the iSAM database to be stored. The user PAD responses are also sent to the python server to fine-tune the affective model component (shown in Figure 3 as the box with a brain icon). We designed three affective computational models – a base model that is simply the emotional data from the IAP database, a random forest model trained on the base model, and a fully-connected model trained on the base model using ResNet50 features. For the base-model, PAD prediction is possible for images already in the IAPs database, so the selection of the final image is only made by inferring an appropriate image category. The random forest and fully-connected models can be trained on updated data sets and can output PAD predictions for any image. In the interest of time, only the fully-connected model has all server endpoints implemented. After fine-tuning, predicted PAD using the updated Machine Learning (ML) model was also sent to the iSAM database to be stored.

III. USER EVALUATION

The iSAM prototype was evaluated with an initial set of four users. We set out to see if iSAM can accurately account for user emotion and if iSAM had any critical design flaws or improvement needs. To accomplish this, we employed a mixed-method approach of task-based evaluation and system log-file analysis. Four university students (1F, 3M, aged 22 to 27) were recruited to play-test iSAM at a home office nearby the university campus. Play-testing sessions lasted from approximately 15-20 minutes, where users were introduced to iSAM by two research evaluators. Users were instructed that they would be helping SAM recover its lost memories by appraising memory gallery photos in virtual reality. The users wore an HTC Vive iVR HMD and were tasked with appraising ten images from the IAPs data-set by utilizing SAM. After the ten images are appraised, the AI model generates a final image based on the subset of the closest matching user reported PAD photos. This image is meant to be closest to the emotion goal state, shown in Table I, where the user then evaluates the final AI curated image. Figure 4 demonstrates one of these four users play-testing iSAM with the described protocol above. After play-testing, we asked the users if they had any additional comments and how they felt about the final image. We then synchronized all recorded data in the iSAM database post hoc for statistical analysis.

IV. RESULTS AND DISCUSSION

Four users playtesting iSAM to each assessed ten training images from IAPs for fine-tuning, in addition to evaluating a final AI curated image. This results in a total of 40 training images and four final evaluation images. Users said they found iSAM to be interesting and found the final image to feel happy – these final images produced by each user

Inference Type	Pleasure	Arousal	Dominance
IAPs Database	1.0776 (E 0.2)	0.1827 (E 0.1)	0.1813 (E 0.2)
Static AI Model	2.5784 (E 0.1)	1.8947 (E 0.14)	0.1315 (E 0.05)
Tuned AI Model	1.9251 (E 0.08)	1.8275 (E 0.15)	0.8520 (E 0.06)
Model Increase	32%	1%	-86%
End Static Model	**3.5634 (E 0.9)**	**1.9167 (E 0.1)**	**1.2302 (E 0.6)**
End Tuned Model	**2.9651 (E 0.9)**	**1.5555 (E 0.2)**	**0.2706 (E 0.5)**
End Model Increase	**17%**	**21%**	**354%**

TABLE II

MEDIAN RESULTS ON USER INFERENCE DIFFERENCE PAD RESPONSES. THE UPPER HALF INDICATES IAPS TRAINING DATA IMAGES (N=40); THE LOWER HALF INDICATES THE FINAL AI CURATED GOOGLE IMAGE AFTER TRAINING DATA PAD RESPONSES (N=4). E INDICATES A STANDARD ERROR; THE SCALE IS IN SELF-ASSESSMENT-MANIKIN FROM 1 TO 9. NOTE A LOWER VALUE CORRESPONDS TO CLOSER ACCURACY TO USER PAD RESPONSE.

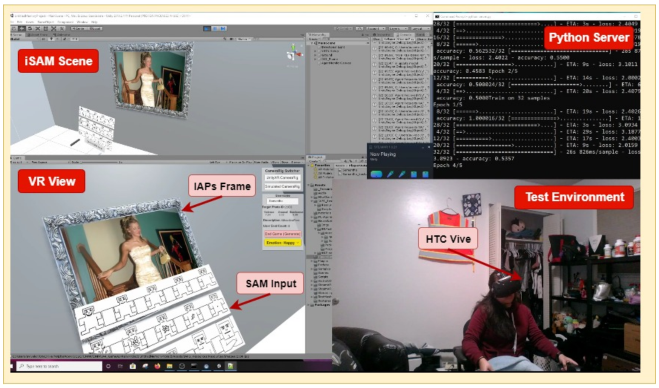

Fig. 4. iSAM Gameplay

can be found in Figure 5. Table II reports the median training image offset between the user's PAD scores for multiple inference types. From the IAPs Database, mean inference was reasonably accurate in predicting user response (best 0.18 offset, worst 1.07 during training images). This indicates consistency with Lang et al's evaluation of IAPs with SAM [13], [24]. In terms of training, the AI Model's tuned performance demonstrated that it learned from the user. The difference in model tuning compared to its static training data showed significant differences after a training session for the final curated image, as demonstrated by the 32%, 1%, and -86% changes from pleasure, arousal, and dominance respectively.

The final generated images were found to be successful in producing a PAD response indicating happiness based on the majority of verbal and PAD responses. The lower half of Table II reports the median final generated images (as shown in Figure 5) and their offset between the user's PAD scores for multiple inference types. The tuned AI model demonstrated a significant percent change from the tuned model on user responses compared to the statically trained model on the IAPs database. These values all displayed a drop in SAM offset, indicating that the tuned model became far more accurate from the users' training responses with significant gains in arousal and dominance predictions.

While the initial results of iSAM were promising, we must

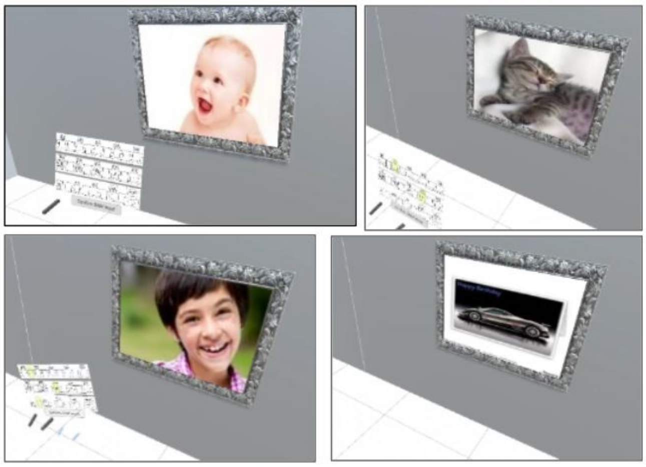

Fig. 5. iSAM's final images produced for each of the four users. These photos indicate optimal PAD values for subjects of babies, children, kittens, and race cars that were chosen by inferring PAD for each individual user from Google image results using AI.

consider some limitations. More users must test iSAM to fully determine its success in predicting and adapting to PAD emotional response. Additionally, more emotional groups beyond happiness should be considered in these playtest sessions. In terms of the AI model, PAD prediction accuracy and tuning can be made more optimal with more training time and with IAPs baseline data enhancement techniques. These limitations are being considered for future studies with our pilot data in mind.

V. CONCLUSION

Through the iSAM prototype, we presented a novel playable experience that employed AI and immersive virtual environments to learn from and adapt to a user's PAD emotional model. We demonstrated a pipeline to enable both users and AI analysis of the International Affective Picture System through a Virtual Reality interface that transported users into a "mind museum" to help SAM recover its lost memories. Our initial play-testing indicated that the AI model was able to improve its PAD emotional prediction for the majority of users through ten training photos from the IAPs and a final curated photo from Google images that was intelligently selected based on the user's PAD responses. Subsequently, this work suggests that it may be possible to bridge runtime emotional models into a virtual environment, which may have substantial implications for the serious games community or any researchers interested in translating emotion models into immersive virtual environments through game engines and AI. For the future, we hope to explore this experience with multi-modal biofeedback to help influence the AI actions and user evaluation. While we feel that this is a step in the right direction, more work must clearly be done to verify the efficacy of iSAM and AI informed experiences that work off of the Pleasure-Arousal-Dominance emotional model. Subsequently, there are far more galleries of the mind to explore and memories to recover in the road ahead.

VI. Acknowledgments

The authors would like to thank the following UC Santa Cruz affiliates for their support and guidance on this project: Conrad Esch, Samantha Conde, Michael Mateas, and Sri Kurniawan.

References

[1] M. Beccue and C. Wheelock, "Research report: Virtual reality for consumer markets," Tractica Research, Tech. Rep., Q4 2016. [Online]. Available: https://www.tractica.com/research/virtual-reality-for-consumer-markets/

[2] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shiban, and A. Mühlberger, "The impact of perception and presence on emotional reactions: a review of research in virtual reality," *Frontiers in psychology*, vol. 6, 2015.

[3] K. R. Lohse, C. G. Hilderman, K. L. Cheung, S. Tatla, and H. M. Van der Loos, "Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy," *PloS one*, vol. 9, no. 3, p. e93318, 2014.

[4] A. Elor, M. Teodorescu, and S. Kurniawan, "Project star catcher: A novel immersive virtual reality experience for upper limb rehabilitation," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 11, no. 4, p. 20, 2018.

[5] A. Elor, S. Kurniawan, and M. Teodorescu, "Towards an immersive virtual reality game for smarter post-stroke rehabilitation," in *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2018, pp. 219–225.

[6] A. Elor, S. Lessard, M. Teodorescu, and S. Kurniawan, "Project butterfly: Synergizing immersive virtual reality with actuated soft exosuit for upper-extremity rehabilitation," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 1448–1456.

[7] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[8] C. Mohn, H. Argstatter, and F.-W. Wilker, "Perception of six basic emotions in music," *Psychology of Music*, vol. 39, no. 4, pp. 503–517, 2011.

[9] C. Collet, E. Vernet-Maury, and A. Dittmar, "Autonomic nervous system response pattern specificity to basic emotions," *International Journal of Psychophysiology*, vol. 1, no. 25, pp. 53–54, 1997.

[10] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.

[11] M. F. Floyd, "Pleasure, arousal, and dominance: Exploring affective determinants of recreation satisfaction," *Leisure Sciences*, vol. 19, no. 2, pp. 83–96, 1997.

[12] J. D. Morris, "Observations: Sam: the self-assessment manikin; an efficient cross-cultural measurement of emotional response," *Journal of advertising research*, vol. 35, no. 6, pp. 63–68, 1995.

[13] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (iaps): Technical manual and affective ratings," *NIMH Center for the Study of Emotion and Attention*, vol. 1, pp. 39–58, 1997.

[14] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[15] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (iaps): Technical manual and affective ratings," *NIMH Center for the Study of Emotion and Attention*, vol. 1, pp. 39–58, 1997.

[16] B. Geethanjali, K. Adalarasu, A. Hemapraba, S. Pravin Kumar, and R. Rajasekeran, "Emotion analysis using sam (self-assessment manikin) scale." *Biomedical Research (0970-938X)*, vol. 28, 2017.

[17] T.-M. Bynion and M. T. Feldner, "Self-assessment manikin," *Encyclopedia of personality and individual differences*, pp. 1–3, 2017.

[18] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik, "The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1643–1652, 2018.

[19] S. B. i Badia, L. V. Quintero, M. S. Cameirao, A. Chirico, S. Triberti, P. Cipresso, and A. Gaggioli, "Towards emotionally-adaptive virtual reality for mental health applications," *IEEE journal of biomedical and health informatics*, 2018.

[20] A. Mehrabian, *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies*. Oelgeschlager, Gunn & Hain Cambridge, MA, 1980.

[21] R. F. Bales, R. F. Bales, R. F. Bales, R. F. Bales, R. F. Bales, T. Adorno, E. Frenkel-Brunswick, D. Levinson, R. Sanford, F. Allport *et al.*, "A comparison between ratings and scorings in the symlog approach." in *Social Interaction Systems: Theory and Measurement*. Farrar and Rinehart New York, 2001, vol. 34, no. 1, pp. 1–82.

[22] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[23] A. Mehrabian, *Nonverbal communication*. Routledge, 2017.

[24] P. Lang and M. M. Bradley, "The international affective picture system (iaps) in the study of emotion and attention," *Handbook of emotion elicitation and assessment*, vol. 29, 2007.